

NASA EOS Science Working Group on Data

Data Access and Usability Workshop

Report

February 26, 2004

Contents

| | |
|----------------------------------------------------------|----|
| Introduction and Background | 1 |
| Acknowledgments | 1 |
| Workshop Overview | 2 |
| Common Threads of Outcomes from the Panels | 3 |
| DAAC Services | 3 |
| User Interfaces | 4 |
| Communication..... | 4 |
| Community-specific Custom Data Sets | 6 |
| Data Timelines..... | 7 |
| Calibration | 7 |
| Panel Reports | 7 |
| I. Climate Researchers Panel Report | 7 |
| II. Earth Science Researchers Panel Report | 8 |
| III. Applications and Operations Users Panel Report..... | 9 |
| IV. Education and Outreach Users Panel Report | 11 |
| Conclusions..... | 13 |
| Appendix A – List of Participants..... | 14 |
| Appendix B – Agenda | 16 |
| Appendix C – Panels..... | 17 |
| Appendix D – List of Issues..... | 18 |
| Appendix E – Organizing Committee | 19 |
| Appendix F – Acronyms | 20 |

Introduction and Background

NASA's Earth Observing System (EOS) Science Working Group on Data (SWGD) recently held a workshop on Data Access and Usability. The workshop included panel discussions by a range of users in a variety of discipline areas who described their objectives, the type of EOS data they used, how they got the data and how usable they found the data. The SWGD was especially interested in any barriers users encountered, how the barriers were overcome, and welcomed suggestions for improvements. The workshop was sponsored by the EOS Project Science Office (PSO).

The workshop was held November 5 and 6, 2003 near NASA's Goddard Space Flight Center (GSFC) in Greenbelt, MD. Close to 100 people attended the workshop (Appendix A) including individual users, EOS instrument and science team representatives, NASA Headquarters and Earth Science Data and Information System (ESDIS) project staff. The EOS project and the science teams were given valuable feedback from the users at the workshop.

The goal of EOS is to develop an understanding and expanding scientific knowledge of the total Earth system and to support environmental policy recommendations. The ESDIS project developed and operates the data and information system called the EOS Data and Information System (EOSDIS), which acquires, processes and distributes EOS data to the user community.

The EOS SWGD originated in 2000 within Terra science teams to address concerns about data production. It is now constituted to represent all EOS instrument and science teams, with additional participation by EOS PSO, Data Active Archive Centers (DAACs), Science Investigator-led Processing Systems (SIPS), NASA Headquarters and Data Users – through DAAC User Working Groups (UWGs) and Science Advisory Panels (SAPs). The SWGD helped fill the need for user participation in determining and prioritizing what is done within EOSDIS. The SWGD's focus has changed over time. Initially the focus was on data acquisition, and then it moved to processing and archiving and most recently on distribution. The group's activities are reported on the web site: <http://swgd.gsfc.nasa.gov>.

Acknowledgments

The material in this document was prepared and reviewed by instrument teams, DAAC representatives, and the panel chairs and assembled as a unified report by the SWGD chair. Credit for primary authorship of the bulk of the original written material goes to Roger King, Jim Collatz, Glen Schuster and Tom Schmugge and Robert Wolfe. Additional comments and revisions were provided by Katalin Kovacs, Vince Salomonson, Graham Bothwell, Daniel Ziskin, Kyle Miller, Marilyn Kaminski and John Dwyer.

Workshop Overview

An introduction to the workshop and a presentation of the workshop goals was made by Robert Wolfe (SWGDD chair, NASA GSFC). A number of speakers provided information about the current state and the future of earth science data access and distribution (see the agenda in Appendix B). A summary of input from the DAAC UWGs was made by Eugene Clothiaux (Langley DAAC UWG chair, Pennsylvania State University). This was a comprehensive summary of access and usability issues gleaned by polling the DAAC advisory panels. Two talks were given on current plans for improving data access. The first was by Steve Fox (EOSDIS Core System (ECS), Raytheon) who talked about the near term improvements in science data access. Robin Pfister (ESDIS) gave the second talk about the EOSDIS Clearing House (ECHO), a system designed to enable a broader set of user type profiles (e.g., clients) to be developed that are tailored to various community needs. A summary of a recent relevant National Research Council (NRC) Panel Report “Government Data Centers; Meeting Increasing Demands” was given by one of the NRC panelists Leo Mark (College of Computing, Georgia Institute of Technology). As part of looking into the future, Chaitan Baru (San Diego Supercomputer Center) gave a vision of the future of earth science data processing in a talk titled “GEON – The Geosciences Network”.

The scope of the workshop was limited to active archive issues; the long-term archive issues were not addressed. These activities were addressed by the SWGDD in a previous report on the long-term archive of EOS data.

At the workshop, representatives from the EOS data user community were divided into four panels: climate researchers, other Earth science researchers, applications/operations users, and education/outreach users (Appendix C). Each panel first participated in a cross-panel discussion during the plenary session that addressed data access and use by that user group. Then each panel and other participants met separately to discuss key barriers to the fullest use of the data, and developed recommendations for future actions for ESDIS and the EOS instrument/science teams.

The panel chairs were selected by the workshop organizing committee (Appendix E) and the panelists were then chosen with the help of the panel chairs. Every attempt was made to get actual data users who could articulate their experiences accessing and using EOS data. As part of the preparatory work, the panelists were asked to focus on a list of issues (Appendix D) prepared by the workshop organizing committee.

Overall, the participants felt the workshop was timely and useful. The data users appreciated both the services already being provided and the improvements that have taken place over the last few years: data being staged more quickly, the addition of on-line access to selected datasets through the DAAC Data Pools, and improvements in the user interface. Everyone realized that there is a fiscal reality and that ESDIS cannot do everything. The participants understood that there is also mission reality to consider: NASA is a research and technology agency, not an operational agency, and so it cannot meet all of the needs of the operational and applications communities. Despite this, it was

clear to the participants that there is a commitment from NASA to work with the operational agencies to help develop useful applications from NASA's research.

In this report we first consider the areas that cut across the various user communities, and then give the recommendations from each panel.

Common Threads of Outcomes from the Panels

There were a number of common threads that cross-cut the various user communities. They fell into the following categories: DAAC services, User Interface, Communication, Community-specific Custom Data Sets, Data Timeliness and Calibration.

DAAC Services

There were three areas of DAAC services that were identified where improvements would help lower the barriers to the use of EOS data: 1) Subsetting, 2) Tools, and 3) Alternative Data Formats.

Subsetting is now offered at many of the DAACs but not for all data sets. When subsetting was offered, the users were satisfied with its capability but felt that a more comprehensive implementation was needed. In addition, a number of users expressed an interest in 'power subsetting' (spatially and/or by parameter) across an entire data set. This capability would allow the user to obtain small spatial subsets over user specified sites for one or more products across the entire data-record of a product. Alternatively, a user could retrieve one parameter across the entire data-record of a product. In either case, the amount of unnecessary data transferred to the user could be greatly reduced. This would be similar to the all-product Moderate Resolution Imaging Spectro-Radiometer (MODIS) subsets over long-term validation sites. For many of the large EOS data sets, this power subsetting is only feasible during a reprocessing campaign. Since these campaigns do not occur very often, coordination with the users is needed before they begin.

Services and tools enabling easier data fusion across instruments and missions are needed. The difficulty of co-locating and inter-comparing datasets from different instruments and sometimes across disciplines was seen as a barrier to the use of the data. This difficulty arises partially because of the choice of different grids by the various instrument and science teams and sometimes because of the different formatting of the data within the HDF-EOS formatted products. An example given was the difficulty of comparing MODIS and Multi-angle Imaging Spectro-Radiometer (MISR) products, which are not only are they stored in different grids, but MISR products are stored in complete orbits while MODIS products are stored in tiles.

Different user communities have specific file formats that they have used historically. Though the use of HDF-EOS (an EOS specific extension of the Hierarchical Data Format (HDF)) is becoming more widespread, it is still seen as a barrier by many users. Services

and tools are needed to make EOS data available in alternative formats such as GeoTIFF, JPEG and netCDF. The HDF and HDF-EOS developers who attended became aware of the large demand for translation to other formats.

User Interfaces

There were four areas of the EOS user interface where improvements would lower the barriers to using the data: 1) Reducing the number of clicks, 2) Deploying portals for different user communities, 3) Enabling script-based queries, and 4) Adding a semantic interface.

People who regularly use the EOS Data Gateway (EDG) interface felt that it has improved significantly over the last few years. There are also a number of DAAC-unique and instrument-unique interfaces that only satisfy portions of the user community. Access to on-line data from the Data Pools was considered a major improvement in data access.

Despite this, it was felt that the EDG could still use some improvements, particularly in the area of reducing the number of “clicks”. This could be done by default settings for some of the standard options, e.g., having ftp pull as the default delivery method.

The different user communities felt that portals need to be tailored to specific user communities. The full deployment of ECHO was anticipated as a way of enabling these custom portals.

A number of power users were interested in script based access to the data. This would allow users to automatically search and retrieve data, eliminating the need to manually search and order the data. Users acknowledged that the potential for abuse of this type of ordering had to be addressed.

A semantic interface for searching for data was discussed as a possible long-term goal. This would allow the user to specify something like “List all the MISR granules that contain hurricanes over the Caribbean.” The users understood that this is still a research area.

Communication

Better communication with the user community would help improve the usability of EOS data. This workshop was an example of improving communication: EOS data users had a chance to learn from how others overcome data access problems by using existing, but not well advertised capabilities.

Many current and potential users now use internet search engines, such as Google¹, to find information on the web. The web crawlers used by these engines are not likely to

¹ The use in this report of specific examples of commercial products, services and formats is not meant as an endorsement.

find EOS datasets. Making EOS data sets visible to these engines has the potential to increase the usage of EOS data.

A number of users felt that email lists and bulletin boards were not being used effectively to notify users about the availability of data and to allow users within the communities to communicate with each other. A moderated email list is an effective way of notifying users about significant events in a data-set's lifetime, e.g., product releases, reprocessing campaigns. An email list in combination with a bulletin board is effective because users are able to find historical information for a particular area of interest. The EOS-tools email list/bulletin board is a good example. Development of mailing lists is difficult because of privacy concerns, but other groups have found ways of providing this type of communication while maintaining the appropriate level of privacy.

The EOSDIS Federation of Earth Science Information Partners (Federation) plays a key role as a place where value is added to standard EOS products to service specific users. More links between the DAACs and the Federation community would connect users to vendors who could help solve their particular problem.

The participants recognized that there already are a large number of tools available for using EOS data. However, users are not necessarily aware that a tool is already available that will solve their particular problem. For instance, many users were unaware that the Data Pools include a tool to help users compare products from various Terra instruments. Better advertising of already-available tools would help make EOS data more usable.

The on-line Data Pools were praised as a good data distribution method. Because of the large size of the EOS data sets, not all data can be made available on the pools. However, it was unclear how the user community influences what is stored in the pool and for how long. A mechanism to allow input from the community in this area would help to make better use of this resource.

In some cases, there are users who need very large amounts of data. The DAACs constantly work to balance the user communities' needs through the most efficient use of the DAACs' resources, such as subscriptions, Data Pools, etc. However, servicing these users may cause delays in making data available to other users because of limited resources. Again, a mechanism that allows community input to help determine the relative priority of these large orders would be useful. The peer-review of large orders is one possible mechanism.

Many commercial off-the-shelf (COTS) tools are now available to support EOS data. However, not all of the tools fully support HDF and HDF-EOS formats. Since the tool vendors are primarily interested in revenue and earnings, one of the best ways to provide incentives for the vendors to provide better support of EOS data is through user community demand for tools that meet their needs. Direct incentives from NASA are another option.

Even though this workshop was considered very useful because it brought a number of EOS user communities together, the participants agreed that additional workshops focused on specific user communities are needed. These workshops would facilitate communication within user communities and be more focused on each community's particular needs.

There should be more involvement from each user community throughout the process of developing future EOS products. This way, the users could start using the data sooner after the launch date. This could also influence some instrument and data-system design decisions. There might be small design changes that, if made early, could greatly improve the usefulness of the products to some user communities.

There is a need for better-organized and more-thorough documentation of data products. Some users use many different types of data and they need to get up the learning curve quickly on each new product type. A more consistent presentation of product documentation across DAACs and products would be very useful.

Community-specific Custom Data Sets

There was consensus that EOS standard data products are primarily oriented toward the Earth science researcher community and satisfy their needs. The other groups felt that the products needed some customization before they can become useful to their specific purposes. For instance, the climate modelers need data at a coarser resolution than many of the products, and many expect areas that have persistent clouds to be filled with an estimated (modeled) value. Several recommendations were made to help address this concern.

ESDIS should work with other organizations such as the Federation (including the ESIPS) to help service specific user communities. In many cases, the standard products can be used as a starting point, but the value-added products make them usable by specific user groups.

The DAACs should provide storage for selected custom (massaged) data sets. There would need to be some criteria for selecting the products to be stored and they would have to conform to metadata and format standards.

A small number of highly focused data sets are needed for the education community (K-12). These products should be provided for each instrument/mission. The DAACs have come to the same conclusion and provide this service for many data sets.

Data persistence is a key concern for all of the communities, but in particular the education community. Education users need to be able to build a lesson plan around a data set and then reuse it for multiple years. The Earth Observer web site was given as an example of custom and persistent data sets that were useful to the education community.

Data sets need to be made available at different spatial resolutions, starting with the resolution of the instrument up to coarser resolutions (up to 0.25 degree or more) for global research.

Data Timeliness

The biggest complaint from the applications and operational communities was the timeliness of the data. Many operational users need products in minutes to hours of acquisition. EOSDIS was designed to provide data within a few days of acquisition. The data loses value quickly in many areas: fire fighting, weather prediction, etc. The MODIS Rapid Fire system was given as an example of how to address this concern. The direct broadcast capability was another feature that users are pleased with. Making calibration and science algorithms freely available to the direct broadcast community as soon as possible after launch helps operational users to quickly build the capability to use the data operationally.

Calibration

Inter-calibration of instruments and datasets is key to the usability of data across instruments and missions. For instance, algorithms that start with surface reflectance can more easily accommodate data from multiple instruments if the instruments are inter-calibrated and accurately geolocated.

Panel Reports

I. Climate Researchers Panel Report

Subsetting

Software tools running at the data provider's site need to be further developed that allow users to access only those data relevant to their needs without having to download huge files and extract the fraction needed. These tools should allow users to subset specific data layers with tiles, time series and spatial extent at the provider's site. Some but not enough of this capability has already been implemented including MODIS Data Gateway spatial subsetting, and the MODIS-Atmospheres time series subsetting for certain gridded products.

Custom Data sets

A large segment of the climate modeling community does not want incomplete data sets, i.e., containing holes where there are missing data segments. They want smoothed continuous data fields of certain higher-level products. However, though the climate modeling community can identify their needs, they do not necessarily understand the characteristics of the satellite data and are not by themselves the ones to implement their needed products. The remote sensing community who has the expertise in the

measurements and product algorithms need to develop these products for and in collaboration with the climate modeling community. This interaction would be two way – the remote sensing community would produce these relevant products based on their knowledge of the characteristics of the measurement and of the needs of the user community, and by using these products the climate modelers would provide evaluation to the producers leading to reprocessing and the evolution of better products.

In addition to the development of community-specific data products there is a need to follow the International Satellite Land Surface Climatology Project (ISLSCP) model of providing multiple, cross-disciplinary community selected data products that are co-registered in time and space. MODIS-Atmospheres Discipline has recently released a subset of their own products and tools with these characteristics but this model should be expanded and applied across all EOS disciplines and relevant measurements for example, smoothed, filled Normalized Difference Vegetation Index (NDVI), solar radiation, optical depth, sea surface temperature, etc., time series (e.g., biweekly) globally at 0.25 degree with common land/sea mask. If the goal is to expand the EOS Data user base in the Climate Research Community rapidly then it must be appreciated that the speed of this expansion is a direct function of the production of these corrected and co-registered data sets.

Archiving

While it has been stated that long term archiving is "out of the scope" of this meeting there is serious concern among the members of the climate research community that the long-term stability of EOS data may be vulnerable to political and agency funding uncertainties. It was suggested that links begin to be established with potential archiving institutions to ensure that these data will be maintained in the long term in useable forms.

II. Earth Science Researchers Panel Report

The goal of the Earth Sciences panel and breakout group was to recommend specific actions that could assist NASA in maximizing the value of its existing and future science data products to this community. It was the belief of the group that the DAACs were doing a good job with their basic job of distributing data but that there are many improvements that could be made to make this more user friendly. These can generally be broken down into two categories: improved documentation and easier delivery.

In terms of documentation, better descriptions of the data should be made available. Most of this already exists but is scattered on different web sites. For example many users are not aware that the Advanced Space-borne Thermal Emission and Reflection Radiometer (ASTER) user's guide has been available through the DAAC for over a year. Another item in this area is the availability of software tools for processing the EOS data. Again these tools are scattered among several different sites.

The improvements suggested with regard to the delivery of data mainly had to do with subsetting the data both spatially and temporally. Spatial subsetting is necessary for those

researchers who are interested in limited areas and find some EOS data sets too large and cumbersome to work with easily (e.g. MODIS tiles). It was also suggested that coherent packages for data from different sensors, e.g. MODIS, MISR, ASTER, etc., be made available for limited areas to facilitate comparison. This would facilitate the inter-calibration of the products from the different sensors.

Temporal subsetting is important for the study of time series of data. Presently it is very difficult to assemble a time series of data for a limited location.

The “classic” barriers identified below in the Applications and Operations Users Panel Report (III) are also appropriate for the Earth Science’s Panel.

III. Applications and Operations Users Panel Report

The goal of the applications and operational users’ panel and breakout group was to recommend specific actions that could assist NASA in maximizing the value of its existing and future science data products to this community. It was the belief of the group that many new applications and operational users could be served at a low marginal cost. This general belief is also the central theme behind the NASA Earth Science Enterprise Applications Division strategic plan (<http://www.earth.nasa.gov/visions/appstrat2002.pdf>). This panel had the largest attendance from the workshop participants, probably a good measure of the desire of this community to use NASA’s observational products.

Several barriers were identified in the breakout group discussion that prevented the fullest use of the data products by this community. However, it is important to note that none of these barriers should be considered as “show-stoppers.” Fundamental to this discussion was NASA’s role as a Science agency and the difficulty this causes for NASA in developing products and services exclusively for the operational community. However, this concern can be overcome through agreements between NASA and the operational community to work together to develop these products and services and then having the DAACs serve them to the community on a timely basis. This, however, may require new operational philosophies for the DAACs as they consider how best to serve the non-science community.

There were several “classic” barriers identified by the community based on operational requirements. Those can be summarized as the absence of the following:

- Continuity of data.
- Consistency in data quality (i.e., cross-calibration, validation of products).
- Timeliness of data delivery (this is going to be application dependent, but one operational user recommended delivery of land products in the same time frame as weather products).
- Data formats that are compatible with the operational community needs (e.g., GIS)

- Many operational agencies use models in their decision-making and this is the basis of the Earth Science Enterprise (ESE) Applications Strategy (Missions to Models to Decision Support). A lack of format translation capability may inhibit the assimilation of imagery and data products into operational (other than NOAA) models.

However, several new barriers were also identified:

- Long lead-time between launch and providing the operational community with data products.
 - The concern here is that by the time the community is served with a product it is too late in the mission life for the operational agencies to provide meaningful feedback to improve the product for their requirements.
- Data fusion will become more important in meeting operational requirements; however, there is a concern that instrument teams do not coordinate well enough to make this possible. The example cited in the workshop was the difficulty in co-locating MISR and MODIS data sets from the same satellite due to their different scanning methods.
 - The following observation was made by the breakout participants - Funding on a project-by-project basis inhibits communication between groups working on system development, the development of common standards, and causes fragmentation of software code which prevents data fusion.
- Operational and application users may need customized products much different than presently available.
- Better bottom-up coordination between providers of data and operational community.
 - It was noted that instruments are designed for research and the application user is often involved late in the process.

Four recommendations, three tactical (short-term) and one strategic (long-term), were formulated in the breakout session. They are prioritized and represent a distilling of many different recommendations into what were deemed as necessary for NASA science data sets to become useful in the broader operational community.

Tactical Recommendation 1

EOSDIS should reexamine how to better serve its many user communities (science, operational, and other distinct communities).

Specific examples:

- Portals tailored to user communities.
- Tools that do simple things with the data (e.g., change detection for the conservation community).

- The “rapid response” model for the forest fire community should be considered an excellent example for replication for other operational requirements.
- Use of other formats (e.g., GeoTIFF, GRIB).

Tactical recommendation 2

Systems and tools that permit image information mining need to be developed. They offer NASA the opportunity to lead in an important emerging area of remote sensing analysis.

Specific examples:

- The archives are rich with information that needs to be exploited to gain new insights for use in science understanding and operational decision support.
- Need tools that can be used for sensor fusion analyses, subsetting, etc.
- Middleware (e.g., ECHO) are good operational tools.

Tactical Recommendation 3

Involve operational users earlier and more closely (i.e., instrument development, product design, and product generation).

Specific examples:

- Have application or mission agency end users on science teams.
- Have operational or application users work with science teams on a day-to-day basis (e.g., Jim Szykman (EPA) at NASA’s Langley Research Center (LaRC)).
- Better understand the requirements of the application user community for sensors, data products, and distribution.

Strategic Recommendation

Heed the lessons learned from the EOS experience in formulating the resource requirements (technical, administrative, and financial) for future missions and products.

IV. Education and Outreach Users Panel Report

NASA satellite data is needed for inclusion in meaningful education products and services. NASA can assist education content developers and teachers to deliver worthwhile data and investigations to the classroom and the public to inspire the Next Generation of Explorers.

To that end schools are accountable for student performance and would embrace simple-to-use imagery supporting real-world investigations and specific learning objectives at elementary, middle and high schools; the higher education communities; and informal education arenas such as museums and science centers.

The education community of learners makes the following three strong recommendations for NASA and Earth data providers in order to fulfill the needs of the education communities:

Data Availability

- a) Provide simple-to-use “current” mission data in formats that can be read simply by browsers (JPEG, GIF, etc.) and widely-used commercial products. To promote student inquiry, mission data should illustrate one or more parameters, and be near real-time or imagery composites over days. Data should be made available routinely and updated automatically for change-over-time studies, including daily, weekly and monthly mean data when possible. Useful imagery includes thoughtful color and image keys.
- b) Provide large higher-resolution data “static” scenes for particular parameters, from which spatial and temporal samples could be extracted and studied. Needs include scenes of various parameters for educators and content developers to ultimately be able to explore local and regional areas of their choice with learners or the public.

Additionally for each a) and b), the following features or points are useful:

- Subsetting [spatial query tools] would be useful, as well as the generation of US and global scenes.
 - Multiple resolutions are necessary and appropriate for different audiences (museums, classroom activities).
 - Persistence issues. Update current data, move imagery and sustain archives. The point was emphasized that Persistence was very important. If you put something on the web for educational purposes, then you must leave it there for a long time.
 - Use of thumbnails should be widespread.
 - Data could be in some additional formats besides JPEG, GeoTIFF, etc., such as ESRI formats.
 - Since NASA uses commercial products (e.g., IDL, MATLAB, etc.), the Strategic Evolution of Earth Science Enterprise Data Systems (SEEDS) Re-use Working Group can help the education community to reuse code resources.
 - Animations should be created when possible (QuickTime or other simple formats used by novice through standard browser tools.)
- c) NASA DAAC archived data should continue to aggressively evolve with the highly regarded Data Pools initiative and the generation of GeoTIFF file formats. DAACs should also include simpler formats such as JPEG and GIF with strong cataloging, and the updating the newest granules. A tool, such as ECHO, should evolve for users to create a useful interface (i.e., “My ECHO”). The interface should utilize thumbnail imagery, and available data for custom, easy-to-access Level 3+ data products for content developers and middleware professionals.

Search Engine Friendliness

Data must be discoverable by internet search engines (e.g. Google). Currently most good data and datasets can not be found. Flat map files with human readable descriptions are essential. Data must be found by location, topic or parameter, and date. Thumbnails are important. In addition, certain available “push” technologies can be adopted to distribute the data for education users. Necessary are low-level descriptions of what the images mean; data should be organized, cross-referenced and searchable according to a topic or phenomenon or parameter (e.g., ozone hole, fires, ice, hurricanes, floods, vegetation, current ocean winds, etc.)

Communication between Science and Education Communities

a) The dialogue and work with educators and science teams must continue to insure that the representation of data is most appropriate for the education level of each grade or public user. Visualizations that draw out classroom lesson objectives at various grade levels (or that are intended for the public) must not be too complex.

b) A special workshop or forum should be held for science teams and educators. This venue will serve both communities. The science teams can share the types of information that come from the knowledge of mission requirements or instrument parameters. The educators can articulate the specific needs that will help incorporate the data into the curriculum. This includes recommendation to the science teams by educators for including essential parameters as well as the educators’ unique spatial and temporal needs regarding the data. Educators can articulate the specific learning goals that will be met utilizing the data.

Conclusions

During a feedback session at the close of the workshop, the participants came to a consensus that the workshop was worthwhile and highly beneficial. There was a proposal that an assessment be done in a year to determine if the workshop recommendations are being adopted. One way to do this would be through the SWGD by a poll of panelists and users. Another way would be to create a set of metrics based on the recommendations that would allow for a balanced assessment of the progress. A number of participants felt that additional user workshops would be useful, with some workshops focused on specific user communities.

The EOS data sets are being used successfully by a broad range of users who have made specific recommendations to help ESDIS focus their resources in order to make EOS data more accessible and usable. ESDIS’s challenge will be to prioritize these recommendations and to address them in the most cost effective manner possible.

Appendix A – List of Participants

| Last | First | Affiliation |
|--------------|--------------|------------------------------------------------------------------------|
| Abrams | Michael | NASA Jet Propulsion Laboratory |
| Arnone | Robert | DOD Naval Research Lab |
| Bambucus | Gyra | NASA Goddard Space Flight Center |
| Bane | Bob | NASA Goddard Space Flight Center Data Usability Group |
| Barkstrom | Bruce | NASA Langley Research Center |
| Baru | Chaitan | San Diego Supercomputer Center |
| Behnke | Jeanne | NASA Goddard Space Flight Center |
| Bethel | Glenn | USDA |
| Bjorn | Eng | NASA Jet Propulsion Laboratory |
| Bock | Christopher | NASA Goddard Space Flight Center |
| Bounoua | Lahouari | NASA Goddard Space Flight Center |
| Bromm-Simonn | Gloria | Massachusetts Institute of Technology |
| Bull | Michael | NASA Jet Propulsion Laboratory |
| Caron | Bruce | The New Media Studio |
| Casey | Kenneth | NOAA National Oceanographic Data Center |
| Clothiaux | Eugene | Pennsylvania State University |
| Collatz | Jim | NASA Goddard Space Flight Center |
| Cuddy | David | NASA Jet Propulsion Laboratory |
| Davis | Paul | University of Maryland |
| Descloitres | Jacques | MODIS Rapid Response |
| Dickinson | Robert | Georgia Institute of Technology |
| Echeverri | Mark | Raytheon ITSS |
| Ferebee | Michelle | NASA Langley Research Center |
| Folk | Mike | University of Illinois/National Center for Supercomputing Applications |
| Fontaine | Kathy | NASA Goddard Space Flight Center |
| Fox | Stephen | Raytheon |
| Friedl | Lawrence | NASA Headquarters |
| Geller | Gary | NASA Jet Propulsion Laboratory |
| Griffin | Vanessa | NASA Goddard Space Flight Center/EOSDIS Science Operations Office |
| Gunther | Fred | Computer Sciences Corporation |
| Hughes | Kent | NOAA/NESDIS/ORA/ORAD |
| Hyer | Edward | University of Maryland Geography |
| Imam | Bisher | University of California, Irvine |
| Iona | Glenn | NASA Goddard Space Flight Center |
| Jones | Dave | StormCenter Communications. Inc. |
| Kalvelage | Tom | Land Processes DAAC |
| Kaminski | Marilyn | National Snow and Ice Data Center |
| Kempler | Steve | NASA Goddard Space Flight Center Earth Sciences DAAC |
| King | Roger | Mississippi State University |
| Knoche | Jeremiah | Oregon State University Geosciences |
| Knopf | William | NASA Headquarters |
| Kovacs | Katalin | SSAI for NASA Goddard |
| Leptoukh | Gregory | NASA Goddard Space Flight Center |
| Lewicki | Scott | NASA Jet Propulsion Laboratory |
| Lillibridge | John | NOAA Laboratory for Satellite Altimetry |
| Little | Michael | NASA Langley Research Center |
| Lowe | Dawn | NASA ESDIS |

SWGD Data Access and Usability Workshop Report

| | | |
|------------|--------------|-------------------------------------------------------------------|
| Lu | Yun-Chi | NASA Goddard Space Flight Center |
| Maiden | Martha | NASA Headquarters |
| Malaret | Erick | Applied Coherent Technology |
| Mark | Leo | College of Computing |
| McCabe | Matthew | Princeton University |
| McDonald | Ken | NASA Goddard Space Flight Center |
| Miller | Kyle | NASA Jet Propulsion Laboratory |
| Moe | Karen | NASA Goddard Space Flight Center |
| Morin | Paul | University of Minnesota Geology and Geophysics |
| Moses | John | NASA Goddard Space Flight Center |
| Murphy | Kevin | NASA ESDIS |
| Olsen | Edward | NASA Jet Propulsion Laboratory |
| Percivall | George | NASA Goddard Space Flight Center |
| Pfister | Robert | NASA Goddard Space Flight Center |
| Pollicelli | Fritz | NASA Goddard Space Flight Center |
| Ramapriyan | H. K. (Rama) | NASA Goddard Space Flight Center |
| Randall | David | Colorado State University |
| Ranson | Jon | NASA Goddard Space Flight Center |
| Raugh | Anne | University of Maryland |
| Reber | Skip | NASA Goddard Space Flight Center |
| Ridgway | Bill | Science Systems & Applications, Inc. |
| Robinson | Jon | Raytheon Information & Infrastructure Systems |
| Salomonson | Vincent | NASA Goddard Space Flight Center Earth Science Directorate |
| Schmaltz | Jeff | NASA Goddard Space Flight Center |
| Schmugge | Thomas | USDA Agricultural Research Service/Hydrology & Remote Sensing Lab |
| Schumacher | Joe | Socioeconomic Data and Applications Center |
| Schuster | Glen | U.S. Satellite Laboratory |
| Smith | Peter | NASA Goddard Space Flight Center DAAC |
| Stanley | Thomas | NASA Stennis Space Center |
| Starrs | Siobhan | Smithsonian Institute |
| Steyaert | Lou | NASA Goddard Space Flight Center |
| Stockli | Reto | NASA Goddard Space Flight Center |
| Strand | Holly | World Wildlife Fund |
| Sun | Guoqing | NASA Goddard Space Flight Center |
| Taber | Michael | University of Northern Colorado |
| Teng | Bill | NASA Goddard Space Flight Center |
| Tyahla | Lori | Global Science & Technology, Inc. |
| Ullman | Richard | NASA ESDIS |
| Vazuqez | Jorge | NASA Jet Propulsion Laboratory |
| Vollmer | Bruce | NASA Goddard Space Flight Center |
| Weinstein | Beth | NASA Goddard Space Flight Center |
| Weitz | Dick | Foundation for Earth Science |
| White | Benjamin | University of Maryland Global Land Cover Facility |
| Whittaker | Tom | University of Wisconsin-Madison SSEC/CIMSS |
| Willard | Ted | Computer Sciences Corporation |
| Williams | David | Environmental Protection Agency |
| Wolfe | Robert | NASA Goddard Space Flight Center |
| Yang | Jingli | Earth Resources Technology, Inc. |
| Yin | Zhangshi | Global Sciences & Technology Inc. |
| Zhao | Peisheng | George Mason University |
| Ziskin | Daniel | NCAR MOPITT |

Appendix B – Agenda

Day 1 (Wednesday, November 05, 2003)

8:00 Registration

8:30 Introduction – *Robert Wolfe*

What are the overriding barriers to fullest use of the EOS data for science and applications - how are we trying to address them in this workshop?

9:00 DAAC User Working Group (UWG) Presentation – *Eugene Clothiaux*

Addressing the questions "What does the UWGs see as the most serious barriers restricting full access and use of EOS data?"

9:45 ESDIS presentation of things to come

- Near Term Science System Data Access Improvements - *Steve Fox (Raytheon)*
- EOSDIS Clearing House (ECHO) - *Robin Pfister (ESDIS)*

10:30 Break

10:45 Presentation on NRC Panel Report that looked into "Government Data Centers; Meeting Increasing Demands" – *NRC Panelist: Leo Mark*

11:15 Climate Researchers Panel discussion – *Chair: Jim Collatz*

12:00 Lunch

1:00 Earth Science Researchers Panel discussion – *Chair: Tom Schmugge*

1:45 Applications and Operations Users Panel discussion – *Chair: Roger King*

2:30 Break

2:45 Education and Outreach Users Panel discussion – *Chair: Glen Schuster*

3:30 Concurrent break-out sessions:

Four groups (climate researchers, earth science researchers, and applications/operations and education/outreach users), where each group comes back with:

- *statements of key barriers to fullest use of the data and information*
- *recommendations for future actions for EOSDIS, SIPS and Instrument Teams*

5:00 Adjourn

Day 2 (Thursday, November 06, 2003)

8:30 Keynote speech: "GEON: The Geosciences Network" – *Chaitan Baru*

9:15 Breakout sessions continued

10:00 Break

10:10 Feedback from breakout sessions – *Panel Chairs (20 min. each)*

11:30 Discussion and summary of recommendations – *Led by Robert Wolfe*

12:00 Feedback from participants on the value of the workshop.

12:15 Adjourn

[From 1:00 to 4:00 Organizing Committee and Panel Chairs met to write up the recommendations.]

Appendix C – Panels

Climate Researchers Panel

Jim Collatz (chair) – NASA/GSFC
Matthew McCabe – Princeton University
Dave Randall – Colorado State University/Atmospheric Sciences
Bisher Imam – University of California, Irvine
Robert Dickinson – Georgia Institute of Technology
Lahouari Bounoua – Univ. of Maryland C.P./Earth Sys. Science Interdisciplinary Center

Earth Science Researchers Panel

Tom Schmugge (chair) – USDA/ARS Hydrology & Remote Sensing Lab
Reto Stöckli – NASA/GSFC
Guoqing Sun – University of Maryland, NASA/GSFC
Thomas Whittaker – University Of Wisconsin-Madison
Ed Hyer – Univ. of Maryland/Geography
Jeremiah Knoche – Oregon State University/Geosciences

Applications/operations Users Panel

Roger King (chair) – Mississippi State University
David Williams – Environmental Protection Agency
Kent Hughes – NOAA/NESDIS/ORAD/ORAD
Ribert Arnone – DOD/Naval Research Laboratory
Glenn Bethel – USDA
Dave Jones – StormCenter Communications, Inc.

Education and Outreach Users Panel

Glen Schuster (chair) – U.S. Satellite Laboratory
Paul Morin – University Of Minnesota
Bruce Caron – The New Media Studio
Siobhan Starrs – Smithsonian Institute
Michael Tabor – University of Northern Colorado/Earth Sciences

Appendix D – List of Issues

Purpose

- The purpose of these issues is to provide a framework for deliberations about how data is obtained and how it can be used — in the most beneficial manner for both the user and the producer
- The list does not cover every possible issue and other relevant topics are welcome
- The list is not an agenda for the meeting but it is hoped that responses can be formulated from the meeting discussions
- The result may be a stand-alone set of responses, or simply a stimulus for developing a set of recommendations

Learning From Where We Are Now

1. In-depth understanding of a range of specific data applications — may be possible in the workshop if the panelists can drill down into the applications on behalf of both themselves and their communities
2. Patterns in the use of data that could be utilized in order to provide a better service and ways this may be realized
3. Barriers that prevent the fullest use of data for the user's applications
4. Other systems in existence that meet user needs better than current NASA facilities and services and what can be learned from those systems
5. Whether products are structured in the best way for users, what is missing, what is extraneous
6. What tool sets are currently available, whether they are sufficiently well developed, and how tool set development might be approached more successfully
7. Current user needs for documentation of data, services, and tools

Evolving Beyond Where We Are

8. Standards or formats that would help usability
9. Potential enhancements to data packaging that make data easier to use
10. Specific new services from DAACs that would make the lives of current users easier
11. Future changes to the approach taken by users over the next five years and how their data usage is expected to change
12. How the needs for data and documentation will evolve with respect to the types of sensors to be operational in 15-20 years

Thinking Differently but Practically

13. Potential for and practicability of retiring something (giving it up) and replacing it with something more useful
14. Two things from each panelist and workshop participant that could be asked of the government or others to improve data usability and access

Appendix E – Organizing Committee

Robert Wolfe - EOS SWGD Chair, NASA/GSFC

Graham Bothwell - MISR Data Processing Lead, NASA/JPL

Jon Ranson - EOS Terra Project Scientist, NASA/GSFC

Daniel Ziskin - MOPITT Data Manager, NCAR

Vanessa Griffin - ESDIS Science Operations Manager, NASA/GSFC

Vince Salomonson - MODIS Science Team Leader, NASA/GSFC

Appendix F – Acronyms

| | |
|------------|-----------------------------------------------------------------------|
| ARS | (USDA) Agricultural Research Service |
| ASTER | Advanced Spaceborne Thermal Emission and Reflection Radiometer |
| COTS | Commercial off-the-shelf |
| DAAC | Distributed Active Archive Center |
| Data Pool | (DAAC) Data Pool (an on-line ftp access to EOS data) |
| DOD | Department of Defense |
| ECHO | EOSDIS Clearing House |
| ECS | EOSDIS Core System |
| EDG | EOS Data Gateway |
| EOS | Earth Observing System |
| EOSDIS | EOS Data and Information System |
| EPA | Environmental Protection Agency |
| ESDIS | Earth Science Data and Information System |
| ESRI | Environmental Systems Research Institute |
| ESE | Earth Science Enterprise |
| Federation | (EOS) Federation of Earth Science Information Partners Organization |
| ftp | File Transfer Protocol |
| GEON | Geosciences Cyberinfrastructure Network |
| GeoTIFF | Geo(graphic) Tagged Image File Format |
| GIF | Graphics Interchange Format |
| Google | (a commercial internet search engine) |
| GRIB | Gridded In Binary (data format) |
| GSFC | (NASA) Goddard Space Flight Center |
| HDF | Hierarchical Data Format |
| HDF-EOS | (an extension of HDF for EOS data) |
| IDL | Interactive Data Language (a commercial remote sensing tool) |
| ISLSCP | International Satellite Land Surface Climatology Project |
| JPEG | Joint Photographic Experts Group (image format) |
| JPL | (NASA) Jet Propulsion Laboratory |
| K-12 | Kindergarten through 12 th Grade |
| LaRC | (NASA) Langley Research Center |
| MATLAB | (a commercial tool for image processing) |
| MD | Maryland |
| MISR | Multi-angle Imaging Spectro-Radiometer |
| MODIS | Moderate Resolution Imaging Spectro-Radiometer |
| MOPITT | Measurements of Pollution in the Troposphere |
| NASA | National Aeronautics and Space Administration |
| NCAR | National Center for Atmospheric Research |
| NDVI | Normalized Difference Vegetation Index |
| NESDIS | (NOAA) National Environmental Satellite, Data and Information Service |
| netCDF | Network Common Data Format |
| NOAA | National Oceanic and Atmospheric Administration |
| NRC | National Research Council |
| ORA | (NESDIS) Office of Research and Applications |

| | |
|-----------|---------------------------------------------------------------------|
| ORAD | (NESDIS) Operational Requirements Area Directorate |
| PSO | (EOS) Project Science Office |
| QuickTime | (a commercial image format) |
| SIPS | (EOS) Science Investigator-led Processing Systems |
| SEEDS | (NASA) Strategic Evolution of Earth Science Enterprise Data Systems |
| SWGD | (EOS) Science Working Group on Data |
| USDA | United States Department of Agriculture |
| UWG | User Working Group |